Draft Guidance for Industry and FDA Staff

Clinical Performance Assessment: Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Approval (PMA) and Premarket Notification [510(k)] Submissions

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.
Document issued on: October 21, 2009

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit written comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. Alternatively, electronic comments may be submitted to http://www.regulations.gov. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft guidance document, contact Nicholas Petrick (OSEL) at 301-796-2563, or by e-mail at Nicholas.Petrick@fda.hhs.gov; or Joyce Whang (ODE) at, 301-796-6516 or by e-mail at Joyce.Whang@fda.hhs.gov.

CDRH
Center for Devices and Radiological Health

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health

Division of Imaging and Applied Mathematics
Office of Science and Engineering Laboratories

Radiological Devices Branch
Division of Reproductive, Abdominal, and Radiological Devices
Office of Device Evaluation

Preface

Additional Copies

Additional copies are available from the Internet at:
http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm187277.htm. You may also send an e-mail request to dsmica@fda.hhs.gov to receive an electronic copy of the guidance or send a fax request to 240-276-3151 to receive a hard copy. Please use the document number (1698) to identify the guidance you are requesting.

Table of Contents

1 # Draft Guidance for Industry and FDA Staff

2

3 # Clinical Performance Assessment:
4 # Considerations for Computer-Assisted
5 # Detection Devices Applied to Radiology
6 # Images and Radiology Device Data -
7 # Premarket Approval (PMA) and Premarket
8 # Notification [510(k)] Submissions

9

10 *This draft guidance, when finalized, will represent the Food and Drug Administration's*
11 *(FDA's) current thinking on this topic. It does not create or confer any rights for or on any*
12 *person and does not operate to bind FDA or the public. You can use an alternative approach if*
13 *the approach satisfies the requirements of the applicable statutes and regulations. If you want*
14 *to discuss an alternative approach, contact the FDA staff responsible for implementing this*
15 *guidance. If you cannot identify the appropriate FDA staff, call the appropriate number listed*
16 *on the title page of this guidance.*

17

18 ## 1.   Introduction

19 This draft guidance document provides recommendations to industry, systems and service
20 providers, consultants, FDA staff, and others regarding clinical performance assessment of
21 computer-assisted detection (CADe[1]) devices applied to radiology images and radiology device
22 data (often referred to as "radiological data" in this document). CADe devices are computerized
23 systems that incorporate pattern recognition and data analysis capabilities (i.e., combine values,
24 measurements, or features extracted from the patient radiological data) intended to identify, mark,
25 highlight, or in any other manner direct attention to portions of an image, or aspects of radiology
26 device data, that may reveal abnormalities during interpretation of patient radiology images or
27 patient radiology device data by the intended user (i.e., a physician or other health care
28 professional), referred to as the "clinician" in this document. In drafting this document, we
29 considered the recommendations on documentation and performance testing for CADe devices
30 made during the public meeting of the Radiological Devices Advisory Panel on March 4-5,
31 2008.[2] This draft guidance is issued for comment purposes only.

---

[1] The use of the acronym CADe for computer-assisted detection may not be a generally
recognized acronym in the community at large. It is used here to identify the specific type of
devices discussed in this document.
[2] http://www.fda.gov/ohrms/dockets/ac/cdrh08.html#radiology

1
2 FDA's guidance documents, including this guidance, do not establish legally enforceable
3 responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should
4 be viewed only as recommendations, unless specific regulatory or statutory requirements are
5 cited. The use of the word *should* in Agency guidances means that something is suggested or
6 recommended, but not required.

7 **The Least Burdensome Approach**

8 This draft guidance document reflects our careful review of what we believe are the relevant
9 issues related to clinical performance studies for CADe devices applied to radiological data
10 and what we believe would be the least burdensome way of addressing these issues. If you
11 have comments on whether there is a less burdensome approach, however, please submit your
12 comments as indicated on the cover of this document.

13 # 2.   Scope

14 This document provides guidance regarding clinical performance assessment studies for CADe
15 devices applied to radiology images and radiology device data. Radiological data include those
16 that are produced during patient examination with ultrasound, radiography, magnetic resonance
17 imaging (MRI), computed tomography (CT), positron emission tomography (PET), etc.[3] As
18 stated above, CADe devices are computerized systems intended to identify, mark, highlight, or in
19 any other manner direct attention to portions of an image, or aspects of radiology device data, that
20 may reveal abnormalities during interpretation of patient radiology images or patient radiology
21 device data, by the clinician.
22
23 By design, a CADe device can be a unique detection scheme specific to only one type of potential
24 abnormality, or a combination or bundle of multiple parallel detection schemes, each one
25 specifically designed to detect one type of potential abnormality revealed in the patient
26 radiological data. Examples of CADe devices that fall within the scope of this draft guidance
27 include:
28   • a CADe algorithm designed to identify and prompt microcalcification clusters and masses
29   on digital mammograms,
30   • a CADe device designed to identify and prompt colonic polyps on CT colonography
31   studies,
32   • a CADe designed to identify and prompt filling defects on thoracic CT examination and,
33   • a CADe designed to identify and prompt brain lesions on head MRI studies.
34
35 This draft guidance does not cover clinical performance assessment studies for CADe devices that
36 are intended for use during intra-operative procedures or for computer-assisted diagnostic devices

---

[3] For any use of a contrast imaging agent, we recommend that you verify that such comports with
the regulation, labeling, and indications of the imaging drugs and devices. You may wish to
consult the draft guidance **New Contrast Imaging Indication Considerations for Devices and
Approved Drug and Biological Products (DRAFT)**
(http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126051.pdf) for new
contrast imaging drugs and devices indications.

1   (CADx) and computer-triage devices, whether marketed as unique devices or bundled with a
2   CADe device that, by itself, may be subject to this draft guidance. Below is further explanation
3   of the CADx and computer-triage devices not covered by this draft guidance:
4
5   •   CADx devices are computerized systems intended to provide information beyond
6   identifying, marking, highlighting, or in any other manner directing attention to portions
7   of an image, or aspects of radiology data, that may reveal abnormalities during
8   interpretation of patient radiology images or patient radiology device data by the clinician.
9   CADx devices include those devices intended to provide an assessment of disease or other
10   conditions in terms of the likelihood of the presence or absence of disease, or devices
11   intended to specify disease type (i.e., specific diagnosis or differential diagnosis), severity,
12   stage, or intervention recommended. An example of such a device would be a computer
13   algorithm designed both to identify and prompt potential microcalcification clusters and
14   masses on digital mammograms and also to provide a probability score to the clinician for
15   each potential lesion as additional information.
16
17   •   Computer-triage devices are computerized systems intended to in any way reduce or
18   eliminate any aspect of clinical care currently provided by a clinician, such as a device for
19   which the output indicates that a subset of patients (i.e., one or more patients in the target
20   population) are normal and therefore do not require interpretation of their radiological
21   data by a clinician. An example of this device is a prescreening computer scheme that
22   identifies patients with normal MRI scans that do not require any review or diagnostic
23   interpretation by a clinician.

24   For any of these types of devices, we recommend that you contact the Agency to inquire about
25   regulatory pathways, regulatory requirements, and recommendations about nonclinical and
26   clinical data.

3.   Rationale

28   This draft guidance makes recommendations as to how you should design and conduct your
29   clinical performance assessment studies (i.e., well-controlled clinical investigations) for your
30   CADe device. These studies may be part of your premarket submission to FDA.[4] The
31   recommendations in this document are meant to guide you as you develop and test your CADe
32   device; they are not meant to specify the full content or type of premarket submission that may be
33   applicable to your device.[5] If you would like the Agency's advice about the classification and the

---

[4] This submission may be a premarket notification (510(k)), an application for premarket approval
(PMA), an application for a product development protocol (PDP), an application for a
humanitarian device exemption (HDE), or an application for an investigational device exemption
(IDE).

[5] A 510(k) submission and a PMA application are the most common submission types for the
CADe devices addressed in this draft guidance. As described in the draft guidance **Computer-
Assisted Detection Devices Applied to Radiology Images and Radiology Device Data -
Premarket Notification [510(k)] Submissions**
(http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm1
87249.htm), some CADe devices are Class II regulated under 21 CFR 892.2050 and require a

1 regulatory requirements that may be applicable to your device, you may submit a request under
2 Section 513(g) of the Federal Food, Drug, and Cosmetic Act (the Act).[6]
3
4 Regardless of the type of premarket submission you are required to submit for your device, we
5 recommend that you request the Agency's review of your protocols prior to initiating your
6 standalone performance assessment and clinical performance assessment studies for your CADe
7 device. To request the Agency's review of your protocols, you may submit a pre-submission to
8 the Agency.
9

10 **4.   Clinical Study Design**

11 The clinical performance assessment of a CADe device is intended to demonstrate the clinical
12 safety and effectiveness of your device for its intended use, when used by the intended user and in
13 accordance with its proposed labeling and instructions.
14
15 As described above in the scope, a CADe device, by design, is intended to identify data that may
16 reveal abnormalities during interpretation of patient images or data by the clinician. There is a
17 complex relationship between the CADe output and the clinician such that clinical performance
18 may depend on a variety of factors that should be considered in any study design including:

19 • timing of CADe application in the interpretive process;

20 • physical characteristics of the CADe mark, i.e., size and shape, type of boundary (e.g.,
21   solid, dashed, circle, isocontour), and proximity of the CADe mark to the abnormality;

22 • user's knowledge of the type of abnormalities that the CADe is designed to mark; and

23 • number of CADe marks.

24
25 Your clinical performance assessment should be well-controlled especially if performed in a
26 laboratory setting (i.e., off site of the clinical arena) to preclude or limit various biases that might
27 impact conclusions on the device safety or effectiveness. Some various types of study designs
28 that may be utilized to assess your CADe device include:
29
30 • A field test or prospective reader study (e.g., randomized controlled trial) that evaluates a
31   device in actual clinical conditions. A field test may not be practical in situations, for
32   example, where there is very low disease prevalence that may necessitate enrollment of an
33   excessively large number of patients.
34 • A retrospective reader study consisting of a retrospective case collection enriched with
35   diseased/abnormal cases is a possible surrogate for a field test.

510(k) while others are Class III and require a PMA. For more information on the various device
classes, see Section 513(a)(1) of the Federal Food, Drug, and Cosmetic Act (the Act) (21 U.S.C.
360c(a)(1)).

[6] Section 513(g) of the Act (21 U.S.C. 360c(g)) provides a means for obtaining the Agency's
views about the classification and the regulatory requirements that may be applicable to your
device.

1   • A stress test is another option for the clinical performance assessment of some CADe
2   devices. A stress test is a retrospective study enriched with patient cases that contain
3   more challenging imaging findings (or other image data) than normally seen in routine
4   clinical practice but that still fall within the device's intended use population (see **Section**
5   **5. Study Population**). Note that the use of sample enrichment will likely alter reader
6   performance in the trial compared with clinical practice because of the differences in
7   disease prevalence (and case difficulty for stress testing) between the trial and clinical
8   practice.
9

10 The clinical performance assessment of CADe devices is typically performed by utilizing a
11 multiple reader multiple case (MRMC) study design, where a set of readers evaluate image data
12 under multiple reading conditions or modalities (e.g., readers unaided versus readers aided by
13 CADe). The MRMC design can be "fully-crossed" whereby all readers independently read all of
14 the cases. This design offers the greatest statistical power for a given number of cases. However,
15 non-fully crossed study designs may be acceptable, for example in prospective studies where
16 interpretations of the same patient data by multiple clinicians may not be feasible.
17

18 Whether you decide on a fully-crossed study design or not, we recommend the use of an MRMC
19 evaluation paradigm to assess the clinical performance of a CADe device using one of the study
20 designs described above. A complete clinical study design protocol should be included in your
21 submission. Pre-specification of the statistical analysis is a key factor for obtaining consistent
22 and convincing scientific evidence. We recommend you provide: [7]

23   • a description of the study design;
24   • a description of how the imaging data are to be collected (e.g., make and model of the
25   imaging device imaging protocol) and the expertise of the person collecting the data (e.g.,
26   x-ray technician)
27   • a copy of the protocol, including the following:
28   o hypothesis to be tested and study endpoints,
29   o plans for checking any assumptions required to validate the tests,
30   o alternative procedures/tests to be used if the required assumptions are not met,
31   o study success criteria that indicate which hypotheses should be met in order for the
32   clinical study to be considered a success,
33   o statistical and clinical justification of the selected case sample size,
34   o statistical and clinical justification of the selected number of readers,
35   o image interpretation methodology and relationship to clinical practice,
36   o randomization methods, and
37   o reader task including rating scale used (see **Section 4**, subsection **Rating Scale**);
38   • the reader qualifications and experience;
39   • a description of the reader training;

---

[7] Precisely what information you should provide to FDA will depend largely on the type of
premarket submission required for your device.

1. • a statistical analysis plan (i.e., endpoints, statistical methods) with description of:
2.   o the process for defining truth (see **Section 6. Reference Standard**),
3.   o the details of the scoring technique used (see **Section 4**, subsection **Scoring**), and
4.   o any results from a pilot study supporting the proposed design.
5.
6. Valid estimation of clinical performance for CADe devices is dependent upon sound study
7. design. Aspects of sound clinical study design should include the following:
8. • study populations (both diseased and normal cases) are appropriately representative of the
9.   intended use population;
10. • study design avoids confounding of the CADe effect, e.g., reading session effects
11. • sample size is sufficient to demonstrate performance claims;
12. • truth definition is appropriate for assessment of performance, and uncertainty in the
13.   reference standard is correctly accounted for in the study analysis, if applicable;
14. • appropriate data cohorts are represented in the data set;
15. • readers are selected such that they are representative of the intended population of clinical
16.   users; and
17. • imaging hardware are selected such that they are consistent with current clinical practice.

Evaluation Paradigm and Study Endpoints

19. Study endpoints should be selected to demonstrate that your CADe device is effective (i.e.,
20. that in a significant portion of the target population, the use of the device for its intended uses
21. and conditions of use, when accompanied by adequate directions for use and warnings against
22. unsafe use, will provide clinically significant results).[8] Selection of the primary and
23. secondary endpoints will depend on the intended use of your device and should be fixed prior
24. to initiating your evaluation. Performance metrics based on the receiver operating
25. characteristic (ROC) curve or variant of ROC (e.g., free-response receiver operating
26. characteristic (FROC) curve or location-specific receiver operating characteristic (LROC)
27. curve), in addition to sensitivity (Se) and specificity (Sp) at a clinical action point will be
28. likely candidates as endpoints. Considering Se/Sp and an ROC based endpoint allows
29. evaluation of the device over the entire range of operating points as well as at the usual cut
30. point a reader would act on in practice. Data collection for both sets of endpoints can be done
31. simultaneously within a single reader study. Sensitivity (Se) is defined as the probability that
32. a test is positive for a population of patients with the disease/condition/abnormality while
33. Specificity (Sp) is defined as the probability that the test is negative for a population of
34. normal patients (i.e., patients without the disease/condition/abnormality). An ROC curve is a
35. plot of all sensitivities at all possible specificities. It is a summary of diagnostic performance
36. of a device or a clinician. An FROC curve is a plot of sensitivity versus the number of false
37. positive marks. FROC metrics summarize diagnostic performance when multiple disease

---

[8] See 21 CFR 860.7(e).

1 sites per patient are accounted for in the analysis. See Wagner, *et* al.[9] and the IRCU Report
2 79 [10] for additional details on these assessment paradigms.
3
4 Various summary performance metrics to assess the effectiveness of the use of your CADe
5 device by readers may be employed (and may vary based on the specific device and clinical
6 indication). Examples of these include:

7 • area, partial area, or any other measures, under ROC curve,

8 • area, partial area, or any other measures, under the FROC curve,

9 • area, partial area, or any other measures, under the LROC curve,

10 • reader Se/Sp (or recall rate[11]) pair, and

11 • reader localization accuracy.
12
13 We recommend the inclusion of lesion-based, patient-based, and any other relevant
14 anatomical or image unit-based measures of performance in the assessment. The selection of
15 lesion-based, patient-based or another unit-based measure of performance as a primary or
16 secondary endpoint will depend on the intended use and the expected impact of the device on
17 clinical practice.
18
19 For study endpoints based on the area under the ROC/FROC/LROC curve or partial area
20 under the ROC/FROC/LROC curve, we recommend that you provide plots of the actual
21 curves along with summary performance information for both parametric and non-parametric
22 analysis approaches when possible. See Gur *et* al.[12] for potential limitations of relying on
23 only one type of ROC analyses. As mentioned above, we also recommend that you include a
24 sensitivity/specificity (or recall rate) endpoint in your analysis when an area-based endpoint is
25 used because it is not always straightforward to translate the magnitude of an area under the
26 curve (AUC) change into the magnitude of change expected in clinical practice. Reporting
27 sensitivity/specificity (or recall rate) may provide additional information for understanding
28 the expected impact of a device on clinical practice.
29
30 We recommend that you describe your statistical evaluation methodology, and provide results
31 including:

32 • overall reader performance;

---

[9] Wagner, R. F., Metz, C. E., and Campbell, G., "Assessment of medical imaging systems and computer aids: A tutorial review," **Acad. Radiol.** 14:723–48, 2007.

[10] ICRU Report 79, "Receiver Operating Characteristic Analysis in Medical Imaging," Vol.8 No.1 (2008), Oxford University Press (ISSN 1473-6691).

[11] Recall rate refers to the percentage of patients (including diseased and non-diseased patients) that are called back or recalled for additional medical assessment.

[12] Gur, D., Bandos, A.I., and Rockette, H.E., "Comparing Areas under Receiver Operating Characteristic Curves: Potential Impact of the Last Experimentally Measured Operating Point," **Radiology** 247:12–15, 2008.

1   •   stratified performance by relevant confounders or effect modifiers (e.g., lesion type,
2   lesion size, lesion location, scanning protocol, imaging hardware, concomitant
3   diseases) (see **Section 5**, **Study Population**); and

4   •   confidence intervals (CIs) that account for reader variability, case variability, and truth
5   variability or other sources of variability when appropriate.

6

7 We recommend that you identify and validate your analysis software.[13] You should provide a
8 reference to the analysis approach used, clarify the software implementation, and specify a
9 version number if appropriate. Certain validated MRMC analysis approaches, examples of
10 which can be found in the literature or obtained online, may be appropriate for your device
11 evaluation depending on its intended use and conditions of use.[14,15] If you plan to write
12 your own analysis software we recommend you submit a copy of the code developed along
13 with your validation data.

14

15 The definitions of a true positive, true negative, false positive, and false negative CADe mark
16 should be consistent with the intended use of the device and the characterization of the
17 reference standard (see **Section 6**, **Reference Standard**).

18 **Control Arm**

19 We recommend you assess the clinical performance of your CADe device relative to a control
20 modality. For PMA submissions, a study control arm that uses conventional clinical
21 interpretation (i.e., interpretation without the CADe device) should generally be the most
22 relevant comparator in CADe performance assessment. For CADe devices intended as
23 second readers, another possible control is double reading by two clinicians. For 510(k)
24 submissions, direct comparison with the predicate CADe device may be useful for
25 establishing substantial equivalence. Other control arms can be valid. We recommend you
26 contact the Agency to discuss your choice of a control arm prior to conducting your clinical
27 study.

---

[13] For more information on MRMC analysis software, see, for example, Obuchowski, N. A., Beiden, S. V., Berbaum, K. S., Hillis, S. L., Ishwaran, H., Song, H. H., and Wagner, R. F., "Multi-reader, multi-case ROC analysis: An empirical comparison of five methods," **Acad. Radiol.** 11: 980–995, 2004.

[14] For MRMC literature references, see, for example: Metz, C. E., "Fundamental ROC analysis," **Handbook of Medical Imaging.** Vol. 1. Physics and Psychophysics. Beutel J, Kundel HL, and VanMetter RL (Eds.) SPIE Press, 751–769, 2000; Wagner, R. F., Metz, C. E., and Campbell, G., "Assessment of medical imaging systems and computer aids: A tutorial review," **Acad. Radiol.** 14:723–48, 2007; Obuchowski, N. A., Beiden, S. V., Berbaum, K. S., Hillis, S. L., Ishwaran, H., Song, H. H., and Wagner, R. F., "Multi-reader, multi-case ROC analysis: An empirical comparison of five methods," **Acad. Radiol.** 11: 980–995, 2004.

[15] For online access to software that analyzes MRMC data based on validated techniques, see, for example: LABMRMC software and general ROC software, The University of Chicago: http://xray.bsd.uchicago.edu/krl/roc_soft6.htm (for either quasi-continuous or categorical data); University of Iowa MRMC software: ftp://perception.radiology.uiowa.edu/PUBLIC (for categorical data); OBUMRM software: http://www.bio.ri.ccf.org/html/obumrm.html.

1
2   The study control arm should utilize the same reading methodology as the device arm and be
3   consistent with clinical practice. The same population of cases, if not the same cases
4   themselves, should be in all study arms to minimize potential bias. For designs that include
5   distinct cases in each study arm, we recommend you provide a description and flow chart
6   demonstrating how patients and readers were randomized into the different arms.
7

Reading Scenarios and Randomization

9   Reading scenarios should be consistent with the intended use of the device. We suggest the
10   following as possible reading scenarios for inclusion as part of the clinical testing:

11
- a conventional reading without the CADe device (i.e., reader alone);

12
- a second-read in which the CADe output is displayed immediately after conducting a
13   conventional interpretation; and

14
- a concurrent or simultaneous read in which the CADe output is available at any time
15   during the interpretation process.

16
17   You should randomize readers, cases, and reading scenarios to reduce bias in performance
18   measures. We recommend you describe your randomization methodology and provide an
19   associated flowchart. One approach to randomization is to make use of the principle of Latin
20   squares. For example, when evaluating both concurrent and second-reader modes with a set
21   of 450 cases, a possible study design may consist of first dividing the cases into three groups
22   of 150 cases, A, B and C. Each group is further divided into subsets of fifty cases, which are
23   read with the same reading scenario. If α, β and γ are the index for the conventional reading,
24   the second-read mode and the concurrent reading mode respectively, then reading scenarios
25   and cases can be assigned as follows:
26

27   Image

Group	Reading Session		
	I	II	III
A(150)(50)	α	β	γ
(50)	β	γ	α
(50)	γ	α	β
B(150)(50)	β	γ	α
(50)	γ	α	β
(50)	α	β	γ
C(150)(50)	γ	α	β
(50)	α	β	γ
(50)	β	γ	α

39
40   If the study enrolled four readers, the example above would result in 600=150x4 readings per
41   group per reading session. The order in which the 150 cases are read should be randomized
42   within each group and reading session. Note that the sample sizes used here are for
43   illustrative purposes only. Generally, the sample sizes needed for clinical studies should be

1 representative of the intended use population. Likewise, this example study design illustrated
2 above is not the only one that could be used to validate the effectiveness of your CADe
3 device.
4
5 In case of multiple reading sessions where the same cases are read multiple times, we
6 recommend that each reading session be separated in time by at least four weeks to avoid
7 memory bias. However, longer time gaps may be advisable. For shorter or longer time gaps
8 between reading sessions, we recommend you provide data supporting your proposed time
9 gaps.

Rating Scale

11 You should use conventional medical interpretation and reporting for lesion location, extent,
12 and patient management. ROC-based endpoints (see **Section 4**, subsection **Evaluation**
13 **Paradigm and Study Endpoints**) may support collecting data with a finer rating scale (e.g., a
14 7-point or 100-point scale) when readers rate the lesion and/or disease status in a patient. We
15 recommend providing training to the readers on the use of the rating scale (see **Section 4**,
16 subsection **Training of Study Participants**).

Scoring

18 We refer to the procedure for determining the correspondence between the reader's
19 interpretation and the truth (e.g., disease status) as the scoring process. The scoring process
20 and the scoring definition are important components in the clinical assessment of a CADe
21 device and should be described. We recommend you describe the process (i.e., rationale,
22 definition, and criteria) for determining whether a reader's interpretation corresponds to the
23 truth status established during the truthing process (see **Section 6. Reference Standard** for
24 information on the truthing process).
25
26 In this document, we describe scoring in terms of the clinical performance assessment. A
27 different type of scoring is used to evaluate device standalone performance which is described
28 in the draft guidance entitled **Computer-Assisted Detection Devices Applied to Radiology**
29 **Images and Radiology Device Data - Premarket Notification [510(k)] Submissions.**[16]
30
31 The scoring process for the clinical studies should be consistent with the abnormalities
32 marked by the CADe and the intended use of your device. The scoring process should be
33 described and fixed prior to initiating your evaluation. In your description of the scoring
34 process, we recommend you indicate whether the scoring is based on:

35   - electronic or non-electronic means;

36   - physical overlap of the boundary, area, or volume of a reader mark in relation to the
37   boundary, area, or volume of reference standard;

---

[16]

http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm18
7249.htm

- relationship of the centroid of a reader mark to the boundary or spatial location of reference standard;

- relationship of the centroid of the reference standard to the boundary or spatial location of a reader mark;

- interpretation by reviewing reader(s); or

- other methods.

For scoring that relies on interpretations by reviewing readers, we recommend you provide the number of readers involved, their qualifications, their levels of experience and expertise, the specific instructions conveyed to them prior to their participation in the scoring process, and any specific criteria used as part of the scoring process. When multiple readers are involved in scoring, you should describe the process by which their interpretations are combined to make an overall scoring determination or how their interpretations are incorporated in the performance evaluation, including how any inconsistencies are addressed.

Training of Study Participants

We recommend you specify instructions and provide training to study participants on the use of the CADe device and the details on how to participate in the clinical study. Training should include a description of the device and instructions for how to use the device. For specialized reading instructions or rules (e.g., rules for changing initial without-CADe interpretation when reviewing the CADe marks), we recommend you justify their clinical relevance according to reading task, clinical workflow, and medical practice.

We also recommend that training be provided to the readers on the use of the rating scale (see **Section 4**, subsection **Rating Scale**), especially if such a rating scale is not generally utilized in clinical practice. Such training helps avoid incorrect or un-interpretable results. We recommend that reader training include rating a representative set of normal and abnormal cases according to the study design methodology, and making use of cases that are not part of the testing database.

5. Study Population

Patient data (i.e., cases) may be collected prospectively or retrospectively based on well-defined inclusion and exclusion criteria. We recommend that you provide the protocol for your case collections. Note that cases collected for your clinical trial should be independent of the cases used during your device development and should be new to the readers participating in the clinical assessment of the device. An acceptable approach for acquiring data is the collection of consecutive cases that are within the inclusion and outside of the exclusion criteria from each participating collection site.

Enrichment with diseased/abnormal cases is permissible for an efficient and less burdensome representative case dataset. You may also enrich the study population with patient cases that contain imaging findings (or other image data) that are challenging to clinicians but that still fall within the device's intended use population. This enrichment is often referred to as stress testing.

1 For example, if assessing a CADe device designed to assist in detecting colon polyps, the study
2 population may be enriched with cases containing small polyps. Enrichment may affect reader
3 performance so the extent of enrichment should be weighed against the introduction of biases into
4 the study design.
5
6 The sample size of the study should be large enough such that the study has adequate power to
7 detect with statistical significance your proposed performance claims. If performance claims are
8 proposed for individual subsets, then the sample sizes for these subsets should be determined
9 accordingly to detect these claims with statistical significance. For formal subset analysis, a pre-
10 specified statistical adjustment for the testing of multiple subsets would be statistically necessary.
11
12 The study population should be representative of the intended use population for your device.
13 Your study dataset should include the full range of diseased/abnormal and normal cases. The
14 study should also contain a sufficient number of cases from important cohorts (e.g., subsets
15 defined by clinically relevant confounders, effect modifiers, and concomitant diseases) such that
16 clinical performance estimates can be obtained for these individual subsets. As stated above,
17 powering these subsets for statistical significance may not be recommended unless specific subset
18 performance claims are being included.
19
20 When describing your study population, we recommend you provide specific information, where
21 appropriate, including:

22 • the patient demographic data (e.g., age, ethnicity, race);

23 • the patient medical history relevant to the CADe application;

24 • the patient disease state and indications for the radiologic test

25 • the conditions of radiologic testing, e.g. technique (including whether the test was
26 performed with/without contrast, contrast type and dose per patient, patient body mass
27 index, radiation exposure, T-weighting for MRI images) and views taken

28 • a description of how the imaging data were collected (e.g., make and model of imaging
29 devices and the imaging protocol) and the expertise of the person collecting the data (e.g.,
30 x-ray technician)

31 • the collection sites;

32 • the processing sites if applicable (e.g., patient data digitization);

33 • the number of cases:

34 o the number of diseased cases

35 o the number of normal cases

36 o methods used to determine disease status, location and extent (see **Section 6.**
37 **Reference Standard**);

38 • the case distributions stratified by relevant confounders or effect modifiers, such as lesion
39 type (e.g., hyperplastic vs. adenomatous colonic polyps), lesion size, lesion location,
40 disease stage, organ characteristics (e.g., breast composition), concomitant diseases,
41 imaging hardware (e.g., makes and models), imaging or scanning protocols, collection
42 sites, and processing sites (if applicable); and

1   • a comparison of the clinical, imaging, and pathologic characteristics of the patient data
2   compared to the target population.

3   **Data Poolability**

4   Premarket approval applications based solely on foreign data and otherwise meeting the
5   criteria for approval may be approved if, among other requirements, the foreign data are
6   applicable to the United States (U.S.) population and U.S. medical practice and the studies
7   have been performed by clinical investigators of recognized competence (21 CFR 814.15).
8   You should justify why non-U.S. data reflects what is expected for a U.S. population with
9   respect to disease occurrence, characteristics, practice of medicine, and clinician competency.
10   In accordance with good clinical study design, you should justify, both statistically and
11   clinically, the poolability of data from multiple sites. We recommend that premarket
12   notification applications follow similar quality data practices with regard to foreign data and
13   data poolability. You are encouraged to contact the Agency if you intend to make use of
14   foreign data as the basis of your premarket submission.
15

16   # 6.   Reference Standard

17   For purposes of this document, the reference standard (also often called the "gold standard" or
18   "ground truth" in the imaging community) for patient data indicates whether the
19   disease/condition/abnormality is present and may include such attributes as the extent or location
20   of the disease/condition/abnormality. We refer to the characterization of the reference standard
21   for the patient, e.g., disease status, as the truthing process.
22

23   We recommend that you provide the rationale for your truthing process and indicate if it is based
24   on:

25   • the output from another device;

26   • an established clinical determination (e.g., biopsy, specific laboratory test);

27   • a follow-up clinical imaging examination;

28   • a follow-up medical examination other than imaging; or

29   • an interpretation by a reviewing clinician(s) (i.e., truther(s)).

30

31   We also recommend that you describe the methodology utilized to make this reference standard
32   determination (e.g., based on pathology or based on a standard of care determination). For
33   truthing that relies on the interpretation by a reviewing clinician(s), we recommend you provide:

34   • the number of truthers involved;

35   • their qualifications;

36   • their levels of experience and expertise;

37   • the instructions conveyed to them prior to participating in the truthing process;

38   • all available clinical information from the patient utilized by the truthers in the
39   identification of disease/condition/abnormality and in the marking of the location and
40   extent of the disease/condition/abnormality; and

1 • any specific criteria used as part of the truthing process.
2

3 When multiple truthers are involved, you should describe the process by which their
4 interpretations are combined to make an overall reference standard determination and how your
5 process accounts for inconsistencies between clinicians participating in the truthing process (truth
6 variability) (see **Section 4**, subsection **Evaluation Paradigm and Study Endpoints**). Note that
7 clinicians participating in the truthing process should not be the same as those who participate in
8 the core clinical performance assessment of the CADe device.
9

10 7.   Reporting

11 Reporting of performance results may be guided by the FDA Guidance entitled **Statistical**
12 **Guidance on Reporting Results from Studies Evaluating Diagnostic Tests; Guidance for**
13 **Industry and FDA Reviewers**.[17] We recommend submitting electronically the data used in any
14 statistical analysis in your study including the following:

15 • patient information,

16 • disease or normal status,

17 • concomitant diseases,

18 • lesion size,

19 • lesion type,

20 • lesion location,

21 • disease stage,

22 • organ characteristics.

23 • imaging hardware,

24 • imaging or scanning protocol,

25 • imaging and data characteristics (e.g., characteristics associated with differences in
26   digitization architectures for a CADe using scanned films),

27 • and statistical analysis.
28

29 For more information on submitting data electronically, please see the FDA white paper entitled
30 **Clinical Data for Premarket Submissions**.[18]
31

---

[17]http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm0
71148.htm
[18]http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/
PremarketSubmissions/ucm136377.htm

8. Postmarket Planning for PMAs

FDA applies the "Total Product Life Cycle (TPLC)" model to promote and protect the public health. Premarket approval (PMA) applications should include a postmarket plan to assess the continued safety, effectiveness, and reliability of an approved device for its intended use.

One potential piece of a postmarket plan is a post-approval study (PAS). FDA may require you to conduct a post-approval study as a condition of approval in a PMA approval order (21 CFR 814.82(a)(2)). A post-approval study is not always necessary as a condition of approval. FDA determines whether one is necessary on a case-by-case basis.

In the event your PMA approval order does require a post-approval study, we suggest that the study population characterization include race, age and target population baselines. FDA recommends that the target population include baselines for prevalence of the abnormality to be detected, as well as current screening method sensitivity, positive predictive value (PPV), specificity, negative predictive value (NPV), biopsy rate, and recall rate. FDA further recommends that you include in your study protocol, at a minimum, the following:

- Radiologist training and experience for those participating in the PAS
- User training with the CADe device
- Adjustments to CADe systems that may occur during the study period
- Types of abnormalities detected
- Type of imaging center
- Consecutive enrollment of subjects
- Study sensitivity, PPV, specificity, NPV, biopsy rate, recall rate, false-negative rate, number of missed abnormalities (may consider evaluation of readings at next exam for comparison of missed abnormalities)
- Area under of curve and/or ROC analysis

FDA will work interactively with you to finalize the postmarket plan and/or any post-approval study protocol prior to approval decisions so that they are ready to implement if the device is approved.

For additional information, please refer to the FDA Guidance entitled **Procedures for Handling Post-Approval Studies Imposed by PMA Order; Guidance for Industry and FDA Staff.**[19]

---

[19] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm070974.htm